

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: September 11, 2002, 12:13:28 ; Search time 39.17 Seconds
(without alignments)
2852.699 Million cell updates/sec

Title: US-09-841-132-190
 Perfect score: 5238
 Sequence: 1 MASWTGGQQMGDRDSSLVPHH.....YEYRDASRGYGLSAGSKVRF 1006

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 5

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 95%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802:*

1:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1980.DAT.*
2:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1981.DAT.*
3:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1982.DAT.*
4:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1983.DAT.*
5:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1984.DAT.*
6:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1985.DAT.*
7:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1986.DAT.*
8:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1987.DAT.*
9:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1988.DAT.*
10:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1989.DAT.*
11:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1990.DAT.*
12:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1991.DAT.*
13:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1992.DAT.*
14:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1993.DAT.*
15:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1994.DAT.*
16:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1995.DAT.*
17:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1996.DAT.*
18:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1997.DAT.*
19:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1998.DAT.*
20:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA1999.DAT.*
21:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA2000.DAT.*
22:	/SIDSL/cgcgdata/hold-genseq/genseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	5238	100.0	1006	22 AAG83207	Protein encoded by
2	5231	99.9	1006	21 AAB13639	C. trachomatis pmp
3	5094	97.3	982	21 AAB13633	C. trachomatis pmp
4	5094	97.3	982	22 AAG83201	Protein encoded by
5	5090	97.2	1012	20 AAY16735	C. trachomatis LGV

ALIGNMENTS

RESULT 1

AAG83207

ID AAG83207 standard; Protein; 1006 AA.

XX
AC

DT 05-SEP-2001 (first entry)

DE Protein encoded by *Chlamydia trachomatis* pmpG gene.

Chlamydia; vaccine; infection; fuslon protein; antigen;
 KW pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;
 KW acute respiratory tract infection; Cap1; CF529; OMCB;
 KW polymorphic membrane protein; pmp; thiol specific antioxidant; TSA.

OS chlamydia trachomatis.

AA
PN
WO200140474-A2

07-JUN-2007

04-DEC-2000: 2000WO-US32919

XX
PR 03-DEC-1999. 99US-045468A

PR 19-APR-2000; 2000US-0556877.
PR 20-JUN-2000; 2000US-0598419
PR

XX
DA /CORT-V CORTVA CORR

XX
PI Probst P Phatja A Skeiky VAW enling en scholar T.

XX
WBT: 2001-374931/39.

XX Chlamydia polypeptides and fuson proteins useful for preventing pelvic
PT inflammatory disease, trachoma, acute respiratory tract infections,
PT atherosclerosis and heart disease -

XX
PS
Claim 2: Page 214-216: 295pp: English:

The present sequence is provided in a specification relating to compounds and methods for the treatment and diagnosis of chlamydial infection. The compounds provided include polypeptides and fusion proteins comprising immunogenic portions of Chlamydia antigens and DNA sequences encoding such polypeptides. They are useful for vaccinating against chlamydial infection, which causes pelvic inflammatory disease, trachoma, acute respiratory tract infections, atherosclerosis and heart disease.

AA	Sequence	1006 AA;
SO	Sequence	1006 AA;

Query Match	100.0%	Score 5238;	DB 22;	Length 1006;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1006;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

QY 1 MASMTGGQQMGGRDSSLVPHHHHHHMI PQGYDGETLTVSPYTVIGPDSGTTVFSAGELT 60
 |||||
 Db 1 masmtggqmqardsslvphhhhhhhmipqgydgetltvsfpytvgdsqatvfsagelt 60
 |||||

QY 61 LKNLDSIAALPI SCF C N L I G S F T V L G R G H S L T E N I R T S T N G A A L S N S A A D G L T T I E G F 1200

QY 121 KEISFSCNCSLLAVLPAAATTKNGSQTPTTTTSTPSNGTIYSKTDLLLLNNEKFSFYSNLVS 180
 . |||||
 ph 121 keisfscncls|lav|naaattknksatofattstosngt|ysktd|llll|nnk|fsf|svn|lv|s 180

QY 181 GDGGAIDAYSLTVQGISKLCVFQNTAQADGGACQVTSFSAMANEAPIAFVANVACVRG 240

Dh 181 gdgaaidaysltvqgisklcvfqnetaqadggacqvtsfsamaneapiafvanvavrg 240

QY 241 GGIAAVQDGGQGVSSSTSTEDPVVSFRNTAVFDGNVARVGGGTSYSGNVAFLLNNGKTL 300

ph 241 qqiaavqdqaaqvssstet edavvs frnt avefdgnvarvvgggtivegnvafllnngktl 300

QY 301 FLNNVAPVYIAAKOPTSGOASNTSNNGDGAIFCKNGAQAGSNSSVSFDEGEGVVF 360
 Db 301 flnnvaspvyiaakoptsgoasntsnnygdgaifckngagaqagsnssvsfdgegvvff 360
 QY 361 SSNVAAGKGGAIYAKKLSVANCQVQFLRNLIANDGGAIYLGESGELSADYGDIIFDGN 420
 Db 361 ssnvaagkggaiyakklsvancqvqflrniandggaiylgesgelsadygdilfdgn 420
 QY 421 LKRTAKENAADVNGVTSSQAISMGSGKITTLLRAKAGHQILFNDPIEMANGNNQPAOSS 480
 Db 421 lkrtakenaadvngvtvssqaismgsgkittllrakaghqilfndpiemangnnqpqass 480
 QY 481 KLLKINDGEGYTDIVFANGSSTLYQNVTTQGRIVLRERAKLSVNSISQTSGLYNEAG 540
 Db 481 kllkindgegytdivfangsstlyqnvttqgrivlreraklsvnsisqtsgslyneag 540
 QY 541 STLDFVTPQPQPPQAAANQLITLSLHLSLSSLLANNAVTPNPPAQDSHPAVIGSTT 600
 Db 541 stldfvtpqpqpqaanqlitlsnlhlsissllannavtpnppaqdshpavigstt 600
 QY 601 AGSVTISGPIFFEDLDOTAYDRYDWLGSNQKINVLKQLGKTPPANAPSDLTIGNEMPKY 660
 Db 601 agsvtisgpiiffedldotaydrydwlgsgnkinkvlqqlgktppanapsdltignemky 660
 QY 661 GYQGSWLAWDPNTANNPVTLKATWTKGYNPGPERVASLVPNSLWGSITLDIRSAHAI 720
 Db 661 gyqgsowlawdpntannpvtlkatwtkgyngpgervaslvpnslwgsitldirshai 720
 QY 721 QASVDGRSYCRGLWVSGVSNFFYHDDRDLGOGYRIISGGYSLGANSYFGSSMFLAFTEV 780
 Db 721 qasvdgrsyrcglwvsgvsnffyhddrdalgogyriisggyslgansyfgssmflaftev 780
 QY 781 FGRSKDYVVCBSNHHACIGSVYLSLQALCGSYLFGDAFTRASYGFGNQHMKTSYTTAE 840
 Db 781 fgrskdyvvcbsnhhacigsvylslqalcg sylfgdaftrasygfgnqhmktsyttae 840
 QY 841 SDVRWNNCLAGETGAGLPVITPESKLYLNLPRFVQAEFSYADHESFTEGDAQAPFS 900
 Db 841 sdvrwnncclagetgaglpv itp esklylnlprfvqae fsyadhesfteegdaqarfs 900
 QY 901 GHLNLSPVGVKFDRCSSSTHPNKYSFMAAYICDAYRTISGETTLLSHOETWTTDAFHL 960
 Db 901 ghlnlspvgvkfdrcssthpnkysfmaayicdayrtisgettllshoetwttdafhl 960
 QY 961 ARHGVVVRGSMYASLTSNIEVYGHGRYEYRDASRGYLSAGSKYRF 1006
 Db 961 arhgvvvrsgsmyasltsnievyghgrye yrdasrgylsagsk yrf 1006

RESULT 2

AAB13639
 ID AAB13639 standard; Protein: 1006 AA.

XX AC AAB13639;

XX DT 02-FEB-2001 (first entry)

XX DE C. trachomatis pmpG gene protein.

XX KW Chlamydial infection; sexually transmitted disease;

KW pelvic inflammatory disease; PID; tubal obstruction; infertility;

KW trachoma; blindness; acute respiratory tract infection;

KW atherosclerosis; coronary heart disease; antibacterial.

XX OS Chlamydia trachomatis.

XX PN WO200034483-A2.

XX PD 15-JUN-2000.

XX PF 08-DEC-1999; 99WO-US29012.

XX

PR 08-DEC-1998; 98US-0208277.
 PR 08-APR-1999; 99US-0288594.
 PR 01-OCT-1999; 99US-0410568.
 PR 22-OCT-1999; 99US-0426571.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;
 XX
 DR WPI; 2000-431303/37.
 XX
 PT Isolated polypeptide for diagnosis and treatment of Chlamydia infection
 PT comprises immunogenic portion of Chlamydia antigen, which comprises
 PT amino acid sequence encoded by polynucleotide sequence
 XX
 PS Claim 2; Pages 208-210; 256pp; English.
 XX
 CC The present invention relates to new nucleic acid sequences and the
 CC proteins encoded by the nucleic acid sequences. The encoded proteins
 CC comprise an immunogenic portion of a Chlamydia antigen. The encoded
 CC proteins are useful for the serodiagnosis and treatment of Chlamydia
 CC infection. Chlamidia are intracellular bacterial pathogens that are
 CC responsible for a wide variety of human infections. C. trachomatis
 CC infection is one of the most common sexually transmitted diseases and can
 CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction
 CC and infertility. Trachoma due to ocular infection with C. trachomatis is
 CC the leading cause of preventable blindness worldwide. C. pneumonia is a
 CC major cause of acute respiratory tract infections in humans and is also
 CC thought to play a role in the pathogenesis of atherosclerosis and
 CC coronary heart disease. The present sequence is a protein isolated in the
 CC present invention.
 XX
 SQ Sequence 1006 AA;
 Query Match 99.9%; Score 5231; DB 21; Length 1006;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1005; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MASMTGGQQGRSSLLVPHHHHHMIPQGIYDGETLTVSPYVIGDPSTTVFSAGELT 60
 Db 1 masmtggqmgrdsslvphhhhhmipqgiydgeltitvspyvigtptvtsagelt 60
 QY 61 LKLNDSIAALPLSCFNGLLGFTVLGRGHSILTFENIRTSNGAALSAAADGLFTIEGF 120
 Db 61 lklndsiaalplscfngllgftvlgrghsltfenirtstngaalsaaadglftiegf 120
 QY 121 KELSFSNCNSLLAVLPAATTNKGSQTPTTSTPSNGTIYSKTDLLLNNEKFSFYSNLVS 180
 Db 121 kelsfscnclsllavlpaaattnkgsqtpttstpsngtiysktdlllnnekfsfynlvs 180
 QY 181 GDGGAIDAKSLTVQGTISKLCVFOENTAAQDGGACQVTSFSAMANEAPIAFVANVAGVRG 240
 Db 181 gdggaidaksltvqgtisklcvfoentaaqdgacqvtsfsamaneapiafvanvagr 240
 QY 241 GGIAAVODGQGVSSSTSTEDPVVFSRNTAFVDFGNVAVRGGIYSYGNVAFLLNGKTL 300
 Db 241 gglaavodgqgvssststedp v v f s r n t a f v d f g n v a r g g i y s y g n v a f l n g k t l 300
 QY 301 FLNNVAPVYIAAKOPTSGOASNTSNNGDGAIFCKNGAQAGSNSSVSFDEGEGVVF 360
 Db 301 flnnvaspvyiaakoptsgoasntsnnygdgaifckngagaqagsnssvsfdgegvvff 360
 QY 361 SSNVAAGKGGAIYAKKLSVANCQVQFLRNLIANDGGAIYLGESGELSADYGDIIFDGN 420
 Db 361 ssnvaagkggaiyakklsvancqvqflrniandggaiylgesgelsadygdilfdgn 420
 QY 421 LKRTAKENAADVNGVTSSQAISMGSGKITTLLRAKAGHQILFNDPIEMANGNNQPAOSS 480
 Db 421 lkrtakenaadvngvtvssqaismgsgkittllrakaghqilfndpiemangnnqpqass 480
 QY 481 KLLKINDGEGYTDIVFANGSSTLYQNVTTQGRIVLRERAKLSVNSISQTSGLYNEAG 540
 Db 481 kllkindgegytdivfangsstlyqnvttqgrivlreraklsvnsisqtsgslyneag 540

Db 661 twtkgynpervvaslvpsnlwsglildirsahsaicqsdvgrscrglwgsvgsnffyh 720
QY 745 DRDALGQGYRIISGGYISLGANSYFGSSMFGLAFTTEVFGRSKDYVVCRSNHACIGSVYLS 804
Db 721 drdalgggyryisggyslgansyfgsmfglaftevfgrskdyvvcrsnhacigsvyls 780
QY 805 TQOALCGSYLFGDAFIRASYGFGNOHMKTSYTFAEESDVRDNNCLAGETICAGLPIVITP 864
Db 781 tqqalcgsyilfgdafirasygfgnqhmktstyfaeesdvrwnncilageigaglpivitp 840
QY 865 SKLYNELRPFVQAEFYADHESFTTEGQDARAFKSGHLLNLSVPVGKFDRCSSSTHPNK 924
Db 841 sklynelrpfvqaefsyadhesfteegdqarafksgllnlsvpvgkfdrcssthpnk 900
QY 925 YSFMAAICDAYRTISGTETLLSHQETWTTDAFHLARHGWWVRGSMYASLTSNIEVYGH 984
Db 901 ysfmaayicdayrtisgtettllshqetwttdafhlarhgvvvrsgmsyasltsnievygh 960
QY 985 GRYEYRDASRGYGLSAGSKVRF 1006
Db 961 gryeyrdasrygylsagskvxf 982

RESULT 4
AAG83201
ID AAG83201 standard; Protein; 982 AA.
XX
AC AAG83201;
XX
DT 05-SEP-2001 (first entry)
XX
DE Protein encoded by Chlamydia trachomatis pmpG gene.
XX
KW Chlamydia; vaccine; infection; fusion protein; antigen;
KW pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;
KW acute respiratory tract infection; CapI; CT529; OMCB;
KW polymorphic membrane protein; pmp; thiol specific antioxidant; TSA.
XX
OS Chlamydia trachomatis.
XX
PN W0200140474-A2.
XX
PD 07-JUN-2001.
XX
PE 04-DEC-2000; 2000WO-US32919.
XX
PR 03-DEC-1999; 9905-0454684.
PR 19-APR-2000; 2000US-0556877.
PR 20-JUN-2000; 2000US-0598419.
XX
XX (CORI-) CORIXA CORP.
PA
XX
PI Probst P, Bhatia A, Skeiky YAW, Fling SP, Scholler J;
XX
XX WPI; 2001-374831/39.
XX
XX Chlamydia polypeptides and fusion proteins useful for preventing pelvic
PT inflammatory disease, trachoma, acute respiratory tract infections,
PT atherosclerosis and heart disease -
XX
XX Claim 2; Page 189-191; 295pp; English.
XX
XX The present sequence is provided in a specification relating to
CC compounds and methods for the treatment and diagnosis of chlamydial
CC infection. The compounds provided include polypeptides and fusion
CC proteins comprising immunogenic portions of Chlamydia antigens
CC and DNA sequences encoding such polypeptides. They are useful for
CC vaccinating against chlamydial infection, which causes pelvic
CC inflammatory disease, trachoma, acute respiratory tract infections,
CC atherosclerosis and heart disease.
XX
XX Sequence 982 AA;
SQ

Query Match 97.3%; Score 5094; DB 22; Length 982;
Best Local Similarity 99.9%; Pred. NO. 0;
Matches 981; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 25 MIPQGIYDGETLVSPFYTIVIGDPSTGTFVTSAGELTLKLNLDNSTIAALPLSCFNGLLCSFT 84
Db 1 mipqgiydgeltvsvfpytvigdpstgttfvtsageltklndnsiaalpplscfngllgsft 60
QY 85 VLGRGHSLTFTENIRTSNTGAALSNSAADGLFTIEGKELSFNSNCNSLLAVLPAATTNKG 144
Db 61 vlgrghsoltftentrtstngaalsnsaadgltiegfkelfsfncnsllavlpaatnkg 120
QY 145 QTPTTSTPSNGIYISKTDLLLLNNEKFSFYSNLVSDGGAIDAKSLTVGIGSKLCVFQE 204
Db 121 qtpttstpsngtiysktdllllnnekfstynlvsgdgaiddaksltvgigsklcvfqe 180
QY 205 NTAQADGGACQVVTFSAMANEAPIAFVANVAGVGGGIAAVDQGGVSSSTSTEDPVV 264
Db 181 ntaqadggacqvvtfsamaneapiafvanvagr99glaavdgqgqvssststedpvv 240
QY 265 SFSRNTAVEFDGNVAVRGGIYSYGNVAFLLNKGKTLFLNNVASPVYIAAKOPTSGQASNT 324
Db 241 sfsrntavefdgnvarvggiysygnvafllnngktrlflnnvaspvylaaakoptsgasnt 300
QY 325 SNYGGCGAIFCKNGAAGSNNSGSVSFDGEGVVFSSNVAAGKGGAIYAKKLSVANC 384
Db 301 snnyggcgai fckngagagannsgsvsfdgegvvffssnvaagkggaiyakklsvancgp 360
QY 385 VQFLRNANDGGAITYLGESGELSLSADYDGIIPDGNLKRRTAKENAADVNGVTVSSQAISM 444
Db 361 vqflrniandggaiyigesgelslsadydgiifdgnlkrktakenaadvngvstvssqaism 420
QY 445 GSGGKITTLRAKAGHQLFNDPIEMANGNNQPAOSSKLLKINDGEGYTGDIVFANGSSTL 504
Db 421 gsggkittlrakaghqlfndpiemangnnqpqagsskllkindgegytgdivfangsstl 480
QY 505 YQNVTEQGRIVLREKAKLSVNSLSQTGGSLYMEAGSTLDFVTPQPPQPPAANQLITLS 564
Db 481 yqnvteqgrivlrekaaklsvnslsqtggslymeagstldfvtpqppqppaanelitls 540
QY 565 NLHLSLSLLANNNAVTPNPQASHPAVIGSTTAGSVTISGPIFFEDLDDTAYDRYD 624
Db 541 nlhlsllsllannnavtpnpqashpavigttagstagsvtisgpiffedldtdaydryd 600
QY 625 WLGSNQKINVLKLOLQTKPPANAPSDTLGNEMPKYQYQGSWKLWDPNTANNNGPYTLKA 684
Db 601 wlgsnqkinvlklqlgtkppanapsdltlgnempkyqygqswklawdpntannngpytlka 660
QY 685 TWTKTGYNPGPERVASLVPNSLWGSILLDIRSAHSAIOASVDGRSYCRGLWVSGVSNFFYH 744
Db 661 twtktgynpgpervvaslvpsnlwsglildirsahsaicqsdvgrscrglwgsvgsnffyh 720
QY 745 DRDALGQGYRIISGGYISLGANSYFGSSMFGLAFTTEVFGRSKDYVVCRSNHACIGSVYLS 804
Db 721 drdalgggyryisggyslgansyfgsmfglaftevfgrskdyvvcrsnhacigsvyls 780
QY 805 TQOALCGSYLFGDAFIRASYGFGNOHMKTSYTFAEESDVRDNNCLAGETICAGLPIVITP 864
Db 781 tqqalcgsyilfgdafirasygfgnqhmktstyfaeesdvrwnncilageigaglpivitp 840
QY 865 SKLYNELRPFVQAEFYADHESFTTEGQDARAFKSGHLLNLSVPVGKFDRCSSSTHPNK 924
Db 841 sklynelrpfvqaefsyadhesfteegdqarafksgllnlsvpvgkfdrcssthpnk 900
QY 925 YSFMAAICDAYRTISGTETLLSHQETWTTDAFHLARHGWWVRGSMYASLTSNIEVYGH 984
Db 901 ysfmaayicdayrtisgtettllshqetwttdafhlarhgvvvrsgmsyasltsnievygh 960
QY 985 GRYEYRDASRGYGLSAGSKVRF 1006
Db 961 gryeyrdasrygylsagskvxf 982

RESULT 5
 AAY16735
 ID AAY16735 standard; Protein; 1012 AA.
 AC AAY16735;
 XX
 DT 21-JUL-1999 (first entry)
 XX
 DE C. trachomatis LGV L2 HMW protein.
 KW Chlamydia; high molecular weight protein; HMW protein; urethritis;
 KW bacterial infection; conjunctivitis; lymphogranuloma venereum; LGV;
 KW cervicitis; epididymitis; endometritis; pelvic inflammatory disease;
 KW PID; salpingitis; tubal occlusion; infertility; cervical cancer;
 KW arteriosclerosis; atherosclerosis.
 XX
 OS Chlamydia trachomatis.
 XX
 PN WO9917741-A1.
 XX
 PD 15-APR-1999.
 XX
 PF 01-OCT-1998; 98WO-US20737.
 XX
 PR 02-OCT-1997; 97US-0942596.
 XX
 PA (ANTE-) ANTEX BIOLOGICS INC.
 XX
 PI Jackson JW, Pace JL;
 XX
 FI WPI: 1999-287659/24.
 DR N-PSDB: AAX60539.
 XX
 PT New Chlamydia protein useful for treating conjunctivitis, urethritis
 PT and cervical cancer
 XX
 PS Claim 4; Fig 3; 141pp; English.
 XX
 CC The invention relates to an isolated Chlamydia species high molecular
 CC weight (HMW) protein having an apparent mol. wt. of 105-115 kD as
 CC determined by SDS-PAGE. The HMW proteins and nucleic acid molecules can
 CC be used for preventing, treating or ameliorating a disorder related to
 CC Chlamydia e.g. bacterial infection, conjunctivitis, urethritis,
 CC lymphogranuloma venereum (LGV), cervicitis, epididymitis, endometritis,
 CC pelvic inflammatory disease (PID), salpingitis, tubal occlusion, cervical
 CC cancer, infertility, arteriosclerosis and atherosclerosis. The products
 CC can also be used for detection and diagnosis. The present sequence
 CC represents a C. trachomatis HMW protein.
 XX
 SQ Sequence 1012 AA;

Query Match 97.2%; Score 5090; DB 20; Length 1012;
 Best Local Similarity 99.7%; Pred. No. 0;
 Matches 979; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 25 MIPQIGYDGETLVSPFTVIGDPSGTTVFSAAGELTLKNDNSIAALPLSCFNLGSGFT 84
 Db :|||||
 Db 31 mvpqgiydgeltvysfpytvgdpsgttvsageltknldnsiaalpiscfgnllgsft 90
 QY 85 VLGRGHSITFTENIRSTNGAALSNSAAGLFTIEGFKELSFNSCNLLAVLPAATTNKG 144
 Db :|||||
 Db 91 vlgrghsltftentstngaalsnsaagdlftiegfkslsfnsncnllavipaattnkg 150
 QY 145 QPTPTTSPNGTIYSKTDLLLNNEKFSFYSLNLSYDGGGATDAKSLTVQGISKLCVFEQ 204
 Db :|||||
 Db 151 qtpttstpsngtiysktdlllnneksfylnsvsgdgaidskltvqgisklcvfe 210
 QY 205 NTAQDGGACQVWTSFSAANEAPITAFVANVAGVRGGGIAAVQDQQGQVSSSTSTEDPV 264
 Db :|||||
 Db 211 ntaqadggacqvwtfsfsaneapiafvanvavrgggiaavqdgqggvssststedpv 270.

QY 265 SFSRNTAVEFDGNVARVGGIYSYGNVAFNLNNGKTLFLNNVASPVYIAAKOPTSGOASNT 324
 Db :|||||
 Db 271 sfsrntavefdgnvarvvggiysygnvafnlngktilflnnvaspyiaakqptsqasnt 330
 QY 325 SNNYGDGGAIFCKNGAQAGSNNSGVSFDEGVEWFFSSNVAAGKGAIIYAKKLSVANGCP 384
 Db :|||||
 Db 331 snnygdggaifckngaqagsnnsqsvsfdegvvffssnvaagkggaiyakklsvancgp 390
 QY 385 VOFLRNIANDGGAIIYLGESGELSADYGDIIIFDNLKRTAKENAADVNGVTYSSQAISM 444
 Db :|||||
 Db 391 vqflrniandggailylgesgelsadygdliifdnlkrtakenaadvngvtvssqaism 450
 QY 445 GSGKITTLLRAKAGHOILFNDPIEMANGNNOQAOSKLLKINDGSGYTGDIIVFANGSSYFL 504
 Db :|||||
 Db 451 gsgkittllrakaghqilfndpiemangnnpagaskllkndgegytgdvifangsstl 510
 QY 505 YONVTIEQGRIVLREKAKLSVNSLSQTGGSLYMEAGSTLDFVTPOPPPOPPAANOLITLS 564
 Db :|||||
 Db 511 yqnvtieqgrivlrekaklsvnsisqtggslymeagstwdftvpqppgqppaanglitis 570
 QY 565 NLHLSLSLLANNVNTNPTNPPAQDSHPAVIGSTTAGSVTISGPIFFEDLDDTAYDRYD 624
 Db :|||||
 Db 571 nlhlsllsllannavntnptnppaqdshpavigsttagsvtisgpiiffedlddtaydryd 630
 QY 625 WLGSNOKINVLKQLGTRKPPANAPSDLTLGNEMPKYQGSWKLAWDPNTANNPVTILKA 684
 Db :|||||
 Db 631 wlgsnokinvlkqlgtkppanapsdltlgnemphygyqgswklawdpntannpvtilka 690
 QY 685 TWTKTGYNPGPERVASLVPNSLWMSILDIRSAHSAIAQASVDGSRGYCRGLWWSGVNFFVYH 744
 Db :|||||
 Db 691 twtktyngpervaslvpsnlwmsildirsaahsaiqasvdgsrgylwvsgvsnffyh 750
 QY 745 DRDALGCGYRYISGGYISLGANSYFGSSMFGLAFTVFGRSKDYVVCRSNHHACISGVILS 804
 Db :|||||
 Db 751 drdalgggyryisggyslgansyfgssmfglaftevfgrskdyvvcrsnhhacisgvils 810
 QY 805 TQQAALCGSYLFGDAPIRASYFGNOHMKTSYTFABESDVRWNNCLAGETIGAGLPITVTP 864
 Db :|||||
 Db 811 tqqaalcgstylfgdafirasyfgnohmktsytfabesdvrwddnncclageigaglpitp 870
 QY 865 SKLYLNELRPFVQAEFSYADHESFTEGGDQARAFKSGHLLNLISVPVGVKFDRCSTHPNK 924
 Db :|||||
 Db 871 sklylneirpfvqaefsyadhesfteegdqarafkshllnlisvpvgvkdrcssthpkn 930
 QY 925 YSFMAAYICDAYRTISGTETTLISHOETWTTDAFHLEHGVVVRGSMVASLTSNIEVYGH 984
 Db :|||||
 Db 931 ysfmaayicdayrtisgtettllshqetwttdafhlnghvvrvgmsyasltsnievygh 990
 QY 985 GRYEYRDASRGYGLSAGSKVRF 1006
 Db :|||||
 Db 991 gryeyrdasrgylsagrvrf 1012

Search completed: September 11, 2002, 12:32:05
 Job time: 1117 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 11, 2002, 12:28:33 ; Search time 25.91 seconds
(without alignments)
3730.832 Million cell updates/sec

Title: US-09-841-132-190
Perfect score: 5238
Sequence: 1 MAMTGGQGMGRDSSLVPHH.....YKRDASRGYGLSAGSKVRF 1006

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 95%
Maximum Match 100%
Listing first 45 summaries

Database: PIR-71
1: pir1:
2: pir2:
3: pir3:
4: pir4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Length	ID	Description
1	4986.5	95.2	1013	2 G71460 probable outer mem

ALIGNMENTS

RESULT 1
G71460
probable outer membrane protein G - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C.Species: Chlamydia trachomatis
C.Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C.Accession: G71460
R.Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
A.Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
A.Reference number: A71570; M01D:99000809
A.Accession: G71460
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-1013 <ARN>
A.Cross-references: GB:AE001360; GB:AE001273; NID:93329342; PIDN:ACG6469.1; PID:9332934
A.Experimental source: GB:AE001360; GB:AE001273; strain UW-3/Cx
C.Genetics:
A.Gene: pmpG

Query Match 95.2%; Score 4986.5; DB 2; Length 1013;
Best Local Similarity 97.4%; Pred. No. 2.3e-294;
Matches 957; Conservative 15; Mismatches 10; Indels 1; Gaps 1;

QY	25	MIPOGIYDGETLTVSPFYTVIGDPSGTTVESAGELTLKLNLDNSIALPLSCFGLNLGSGFT	84
DB	31	MIPOGIYDGETLTVSPFYTVIGDPSGTTVESAGELTLKLNLDNSIALPLSCFGLNLGSGFT	90
QY	85	VLGRHSILTEENITSTNGAALSNSAADLFTIEGFEKELSFNSCNLSLAVLPAATTNGKS	144
DB	91	VLGRHSILTEENITSTNGAALSNSAADLFTIEGFEKELSFNSCNLSLAVLPAATTNGKS	150
QY	145	QTPPTTSPNSGTYSTDTLLLNNEKFSFNSLVSDGGAIDAKSLTVGGSILCVQOE	204
DB	151	QTPPTTSPNSGTYSTDTLLLNNEKFSFNSLVSDGGAIDAKSLTVGGSILCVQOE	210
QY	205	NTAAGDAGACQVYTSFSAMANEAPIAFVAVYAGRGGIAAODGQOQGVSSSTEDPPV	264
DB	211	NTAAGDAGACQVYTSFSAMANEAPIAFVAVYAGRGGIAAODGQOQGVSSSTEDPPV	270
QY	265	SFSRNTAVFEEDGNVAVRGGIIYSTGVNAFLNNGKTLFLNNVASFVYIAAOPITSGAASNT	324
DB	271	SFSRNTAVFEEDGNVAVRGGIIYSTGVNAFLNNGKTLFLNNVASFVYIAAOPITSGAASNT	330
QY	325	SNNYGDGGAIFCKNGAQ-AGSNNSGSVFQEGVYFFSSNVAAGKGAIAAKLSVANCG	383
DB	331	SDNYGDGGAIFCKNGAQ-AGSNNSGSVFQEGVYFFSSNVAAGKGAIAAKLSVANCG	390
QY	384	PVOFLRNIAIDGAIYGESEGLSLSDADYGDIFDGNLKTAKENADVNGVYSSOAS	443
DB	391	PVOFLRNIAIDGAIYGESEGLSLSDADYGDIFDGNLKTAKENADVNGVYSSOAS	450
QY	444	MSGGKITTLRAKAGHILFNPIEMANGNPOAOSKLLKINDGEGYTDIYFANCSST	503
DB	451	MSGGKITTLRAKAGHILFNPIEMANGNPOAOSKLLKINDGEGYTDIYFANCSST	510
QY	504	LYQWVTEEGRIYAREAKLSVNSLSTGGSILWEASTLDFYTPPOPPAANOILTL	563
DB	511	LYQWVTEEGRIYAREAKLSVNSLSTGGSILWEASTLDFYTPPOPPAANOILTL	570
QY	564	SNLHLSLSLANNVANNPTNPPAODSHAVIGSTAGSVTISGPIFFEDLDPTAVDRT	623
DB	571	SNLHLSLSLANNVANNPTNPPAODSHAVIGSTAGSVTISGPIFFEDLDPTAVDRT	630
QY	624	DWLGSNOKIVLKLQIGTKPPANAPSDLTIGENPKYGYGSMKLANDPMTANNPYTLK	683
DB	631	DWLGSNOKIVLKLQIGTKPPANAPSDLTIGENPKYGYGSMKLANDPMTANNPYTLK	690
QY	684	ATWTKGTGNPGRVAVSLVNSLMSGLDILRSASAIQAASVDGSRYSRGILVSGVSNFFY	743
DB	691	ATWTKGTGNPGRVAVSLVNSLMSGLDILRSASAIQAASVDGSRYSRGILVSGVSNFFY	750
QY	744	HDRDALGQGYRTISGYSILGANSYFGSSMGFLAFTEYFGRSKDYVYCRSNHHACIGSVYL	803
DB	751	HDRDALGQGYRTISGYSILGANSYFGSSMGFLAFTEYFGRSKDYVYCRSNHHACIGSVYL	810
QY	804	STOALGCGSYLFGDAFIRASVFGNOMKTSYTPAEBSVDVWMDNCLAGEIGLPIVIT	863
DB	811	STOALGCGSYLFGDAFIRASVFGNOMKTSYTPAEBSVDVWMDNCLAGEIGLPIVIT	870
QY	864	PSKLYLNELPFVQAESEYSDHESFTEEGQAARFSGHLNLSVPGVAFDRCSSHPN	923
DB	871	PSKLYLNELPFVQAESEYSDHESFTEEGQAARFSGHLNLSVPGVAFDRCSSHPN	930
QY	924	KYSFMAVYICDAYRTISGTTLLSHOETWTTAPFLHARGVAVRGSMAVSLNSIEVYG	983
DB	931	KYSFMAVYICDAYRTISGTTLLSHOETWTTAPFLHARGVAVRGSMAVSLNSIEVYG	990
QY	984	HGRYERDASRGYGLSAGSKVRF 1006	
DB	991	HGRYERDASRGYGLSAGSKVRF 1013	

Search completed: September 11, 2002, 12:32:59
Job time: 266 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 11, 2002, 12:32:28 ; Search time 15.1 seconds
(without alignments)
2579.594 Million cell updates/sec

Title: US-09-841-132-190
Perfect score: 5238
Sequence: 1 MASMTGGQGMGRDSSSLVPHH.....YEYRDASRGYGLSAGSKVRF 1006

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 95%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4986.5	95.2	1013	1 PMPG_CHLTR	084879 chlamydia t

ALIGNMENTS

RESULT	1
PMPG_CHLTR	STANDARD; PRT; 1013 AA.
ID	084879;
AC	16-OCT-2001 (Rel. 40, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Probable outer membrane protein pmpg precursor (Polymorphic membrane protein G).
DE	PMPG OR Ct871.
GN	Chlamydia trachomatis.
OC	Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.
OX	NCBI_TaxID=813;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=D/UN-3/CX;
RX	MEDLINE=99000809; PubMed=9784136;
RA	Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L., Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V., Davis R.W.;
RA	"Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis."
RT	Science 282:754-759(1998).
RL	-1- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES) (POTENTIAL).
CC	-1- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC	
CC	

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: AE001360; AAC68469.1;
DR InterPro: IPR003368; DUF145.
DR InterPro: IPR003357; OMP.
DR Pfam: PF02415; DUF145; 1.
DR Pfam: PF02385; OMP; 1.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 1013 PROBABLE OUTER MEMBRANE PROTEIN PMPG.
SQ SEQUENCE 1013 AA; 107366 MW; F0927743C0A651DD CRC64;

Query Match 95.2%; Score 4986.5; DB 1; Length 1013;
Best Local Similarity 97.4%; Pred. No. 4.7e-286;
Matches 957; Conservative 15; Mismatches 10; Indels 1; Gaps 1;

QY	25	MIPOGIYDGETLVSPYTVIGDPSTGVFSAGELTLKNDNSIAALPLSCFNLLGSFT	84
DB	31	MIPOGIYDGETLVSPYTVIGDPSTGVFSAGELTLKNDNSIAALPLSCFNLLGSFT	90
QY	85	VLGRGHSFLTFENIRSTNGAALSNSAAGLFTIEGKELSFSCNSLLAVLPAAATNKG	144
DB	91	VLGRGHSFLTFENIRSTNGAALSNSAAGLFTIEGKELSFSCNSLLAVLPAAATNNG	150
QY	145	QTPTTSTPSNGIYKSTDLNLLNNEKFSYSLVSDGGAIDAKSLTVOGISKLCVFOE	204
DB	151	QTPTTSTPSNGIYKSTDLNLLNNEKFSYSLVSDGGAIDAKSLTVOGISKLCVFOE	210
QY	205	NTAQADGGACQVTSFSAMANEAPIAVNAVVGSGGIAAVDGGGVSSSTSTEDPVV	264
DB	211	NTAQADGGACQVTSFSAMANEAPIAVNAVVGSGGIAAVDGGGVSSSTSTEDPVV	270
QY	265	SFSRNTAVEFDGNVARGGGIYSGNVAFNLNCKTLFLNNVAPVIAAKQPTSGOASNT	324
DB	271	SFSRNTAVEFDGNVARGGGIYSGNVAFNLNCKTLFLNNVAPVIAAKQPTSGOASNT	330
QY	325	SNNYGDGAIFCKNGAQ-AGSNNSGVSFDGEGVFFSSNVAAGKGGIYAKKLSVANC	383
DB	331	SDNYGDGAIFCKNGAAGSNNSGVSFDGEGVFFSSNVAAGKGGIYAKKLSVANC	390
QY	384	PVQFLNIANDGGAIYLGESGELSADYGDIIIFDGNLKRKTAKENAAADVNGVTSSQA	443
DB	391	PVQFLNIANDGGAIYLGESGELSADYGDIIIFDGNLKRKTAKENAAADVNGVTSSQA	450
QY	444	MGSGGKITTLRAKAGHOILFNDPIEMANGNNQPAQSSKLLKINDGEGYTGDIIVFANG	503
DB	451	MGSGGKITTLRAKAGHOILFNDPIEMANGNNQPAQSSKLLKINDGEGYTGDIIVFANG	510
QY	504	LYQNVTEQGRIVLREKAKLSVNSLSQTGSLYMEAGSTLDFVTPPQPPQPPAANQLITL	563
DB	511	LYQNVTEQGRIVLREKAKLSVNSLSQTGSLYMEAGSTLDFVTPPQPPQPPAANQLITL	570
QY	564	SNLHLSLLANNVNTNPPNPAQDSHPAVIGSTTAGSVTISGPIFFEDLDDTAYDRI	623
DB	571	SNLHLSLLANNVNTNPPNPAQDSHPAVIGSTTAGSVTISGPIFFEDLDDTAYDRI	630
QY	624	DWLGSNQKINVKLQJLQTKRPPANAPSDLTLCNEMPKYGYGSGKWLAWDPNTANGPYTLK	683
DB	631	DWLGSNQKIDVLQJLQTKRPPANAPSDLTLCNEMPKYGYGSGKWLAWDPNTANGPYTLK	690
QY	684	ATWTKGYNPGPERVASLPVNSLWGSTLDIRSAHSAIQASVDGSRVCRGLWVGVSNNFY	743
DB	691	ATWTKGYNPGPERVASLPVNSLWGSTLDIRSAHSAIQASVDGSRVCRGLWVGVSNNFY	750
QY	744	HDRDALGQGYRIISGGYSLGANSYFGSSMFLAFTEVFGRSKDYVVCRRSHHACISVYL	803

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 11, 2002, 12:32:08 ; Search time 40.86 Seconds
(without alignments)
4259.249 Million cell updates/sec

Title: US-09-841-132-190
Perfect score: 5238
Sequence: 1 MASMTGGQMGHDSLSVPH.....YERYDASRGYGLSAGSKVRF 1006

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 95%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	ID	Description
No matches found			

Search completed: September 11, 2002, 12:36:59
Job time: 291 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: September 11, 2002, 12:25:13 ; Search time 17.9 Seconds
(without alignments)
1372.746 Million cell updates/sec

Title: US-09-841-132-190
Perfect score: 5238
Sequence: 1 MASMTGQQMGRRDSSLVPHH.....YEYRDASRGYGLSAGSKVRF 1006

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 95%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query				
No.	Score	Match	Length	DB	ID
Description					

No matches found

Search completed: September 11, 2002, 12:32:25
Job time: 432 sec

